ELSEVIER

Biophysical Chemistry 123 (2006) 25–28

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

The improvement of the algorithm for order parameter calculation (S^2) from molecular dynamics simulation using the correlation motion function

V.M. Dubyna *, D.B. Kovalskyy, O.S. Ivanova, A.I. Kornelyuk

Institute of Molecular Biology and Genetics, 150 Akad. Zabolotny Str., Kyiv-143, 03143, Ukraine

Received 24 August 2005; received in revised form 3 March 2006; accepted 6 March 2006
Available online 10 March 2006

Abstract

The generalized order parameter, S^2, calculated from MD simulation trajectory using time-dependent internal Correlation Motion Function (CMF) agrees well with NMR derived S^2 processed with the extended model-free analysis approach. However, the former lies considerably lower comparing to simple model-free derived data from NMR experiments. In the present study we analyze possible reasons of such disagreement. In the general case we propose to use preexponential factors from expression for internal CMF rather than ordinary S^2 values. Particularly, in case of the simple model-free S^2 experimental values we suggest comparing them with $S^2_{\text{eff}} = 1 + S^2 - S^2_{\text{f}}$ computed from MD simulation data. We show that the S^2_{eff} values are in a good agreement with NMR derived S^2 values obtained using the simple model-free analysis.

Keywords: Internal correlated motion function; S^2 order parameter; MD simulation; The NMR relaxation data; Extended model-free analysis; Simple model-free analysis

1. Introduction

Adequate and accurate cross comparison of the NMR and MD simulation data is of crucial importance in versatile studies conformational dynamics of proteins. The generalized order parameter, S^2 [1], extracted from NMR [2] and MD data [3–5], is the appropriate indicator of protein backbone motions in computationally feasible timescales. However, the S^2 values extracted from a MD simulation trajectory coincide well with NMR derived S^2 values processed with the extended model-free approach [3], but may lie considerably lower than those from the simple model-free analysis [4]. Such disagreement is usually fixed using artificial adjacent variables such as "window range" (see for example [5]).

In the present study we analyze possible reasons of this disagreement and propose the way to avoid it.

2. Materials and methods

A peptide N–H bond motions are described from the NMR relaxation data using the correlation motion function (CMF) which is expressed for each residue as [1]:

$$C(\tau) = < \mu(t)\mu(\tau + t) >_t \quad (1)$$

where $\mu(t)$ is the unit vector along the N–H bond direction at the time t. When the overall protein motion is not taken into account (i.e. when the coordinate system is associated with protein) the function is called internal correlation motion function (internal CMF). During the NMR data processing the external protein motion is removed [1]. Thus, in our study we will consider the internal correlation motion function only.

On the basis of the GROMACS [6] template a program for correlation motion function calculation was written. As the model system we chose molecular dynamics of HIV-1 protease in water at 310 K during 7.36 ns that was previously computed in our laboratory [4]. Using the program we computed the time-dependent internal CMF for all residues of the HIV-1 protease. The averaging was performed over all pairs of frames obtained in MD simulation. In order to treat internal motions only, the

* Corresponding author.
E-mail address: vdubyna@univ.kiev.ua (V.M. Dubyna).

0301-4622/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2006.03.001

overall translational and rotational movements of HIV-1 protease were removed by least-square fitting of C_α atoms to the initial coordinates. A few typical time dependences of internal CMF are shown in Figs. 1–6 of Supplementary materials.

3. Results and discussion

To interpret the NMR results the internal CMF with respect to time can be approximated either by the simple model-free approach [7]:

$$C(t) = A\exp(-t/\tau) + S^2 \quad (2)$$

or by the extended model-free approach [7]:

$$C(t) = A_f\exp(-t/\tau_f) + A_s\exp(-t/\tau_s) + S^2, \quad (3)$$

where $A=1-S^2$, $A_f=1-S_f^2$, $A_s=S_f^2-S^2$; S^2 is the generalized order parameter, S_f^2 is the order parameter for fast motions, τ_f and τ_s are the correlation times for the fast and slow motions, respectively. The function (2) is characterized by the exponential decay $t\sim\tau$ and is close to a constant at $t>\tau$. The function (3) is characterized by the rapid exponential decay at $t\sim\tau_f$ (if $\tau_f<\tau_s$), by the slower exponential decay at $t\sim\tau_s$ and by almost constant value at $t>\tau_s$. Taking into account these facts and the shapes of the calculated internal CMF (see Supplementary materials), we suggested that with few exceptions (for example, residue Ile50) the internal CMF function could be smoothened by Eq. (3) with $\tau_f<1$ ps. Eq. (2) is not valid for the approximation of the CMF with a fast decay at $t<1$ ps and a slow decay at higher t values. For this reason we smoothened the internal CMF using Eq. (3).

The smoothing was performed using the least-square method. For this purpose the global minimum of the following function was found:

$$L(S^2, S_f^2, \tau_f, \tau_s) = \sum_i P_i(C_i - (S^2 + (1-S_f^2)\exp(-t_i/\tau_f) + (S_f^2-S^2)\exp(-t_i/\tau_s)))^2 \quad (4)$$

where i — number of CMF values calculated from dynamics. C_i is the internal CMF value calculated from the MD simulation data. P_i is the statistical weight of i — the point which is equal to a number of frame pairs with time difference equal to t_i. Our model protein (unbound HIV-1 protease) is a C_2-symmetrical homodimer. The summation in the formula (4) was performed over all calculated CMF values for each respective residue pair in both subunits. In order to minimize function (4) we found the argument values at which all partial derivatives of (4) would be equal to zero. For this purpose the system of four equations (four derivatives with respect to each argument of function (4)) was solved using the Newton method. The minimization was considered to be completed if, either, the modifications of S^2 and S_f^2 values were less than 10% of S^2 and S_f^2 values itself, correspondingly and the modifications of τ values were less than 20% of the corresponding τ values, or if the number of optimization steps was equal to 100. The S^2 and S_f^2 values converge very rapidly comparing to τ values so error for S^2 and S_f^2 estimation during optimization is much lower than 10%. In many cases about 10 steps of optimization are enough to evaluate S and τ values with mentioned accuracy. On the other hand, sometimes the function's local minimum cannot be reached at the non-linear optimization. To avoid endless calculation we restricted the maximal number of steps. The non-linear optimization requires the initial values for variables to be optimized. Generally the function (4) may have a few local minima. We used several initial values for each function (4) argument: 0.1, 0.5, 0.9 (for S^2 and S_f^2); 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100 ps (for τ_f); 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000 ps (for τ_s). For each pattern of these initial values the equation system (see above) was solved and afterwards the arguments set corresponding to the least function (4) value was chosen. In this way we obtained S^2, S_f^2, τ_s and τ_f from the MD simulation data for each residue.

Good coincidences of the CMF graphs with approximated curves (see Supplementary materials) suggest that the expression (3) fits time dependent internal CMF while using S^2, S_f^2, τ_s and τ_f that were evaluated here. The calculated S^2 values (Fig. 1a) are in a good agreement with corresponding NMR derived

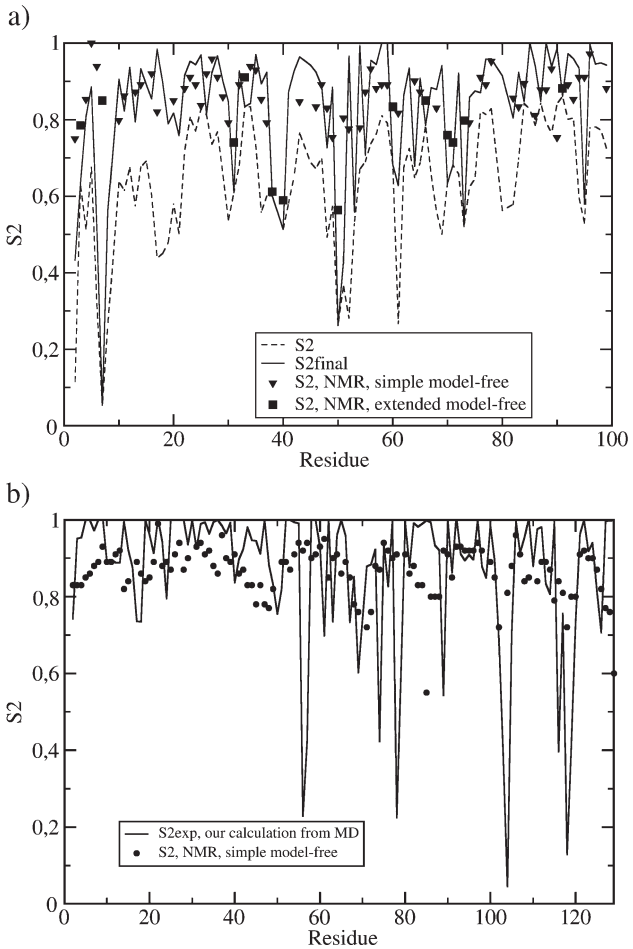

Fig 1. (a) The calculated S^2 and S^2_{final} (equal to S^2 in the case of extended model-free analysis or to S^2_{exp} in the case of simple model-free analysis) values from MD simulation in comparison to experimental data [2] for HIV-1 protease; (b) calculated S^2_{exp} values and experimental results [8] (single model-free analysis) for hen lysozyme.

values obtained with the extended model-free approach [2]. For those residues that were processed with simple model-free approach in NMR experiment our S^2 estimates lay considerably lower. We suggested the reasonable overcome.

The S^2 and τ values are calculated from the NMR relaxation data using the spectral density function [1,7], which can be expressed as

$$J(\omega_i) = \frac{A\tau^2}{(1+\omega_i^2\tau^2)} + \frac{S^2\tau_R^2}{(1+\omega_i^2\tau_R^2)} \quad (5)$$

in the case of simple model-free analysis, and as

$$J(\omega_i) = \frac{A_f\tau_f}{(1+\omega_i^2\tau_f^2)} + \frac{A_s\tau_s}{(1+\omega_i^2\tau_s^2)} + \frac{S^2\tau_R}{(1+\omega_i^2\tau_R^2)} \quad (6)$$

in the case of extended model-free analysis. In these equations τ_f, τ_s and τ_R are fast, slow and overall rotation correlation times. The A, A_f and A_s are preexponential factors from expressions (2) and (3) corresponding terms containing correlation times τ, τ_f and τ_s. The A, A_f and A_s values also may be considered as magnitudes of orientational motions with corresponding correlation time (on corresponding time scale). The S^2 parameter also may be considered as overall (not internal) CMF decay caused by overall diffuse rotation (overall CMF at high t values is closed to 0). It can be seen that each term in Eqs. (2), (3), (5) and (6) contains only parameters for motion in corresponding timescale. On the basis of these facts we suppose that either CMF (magnitude of N–H bond orientation disorder) or spectral density function additively includes contribution of motions in different timescales and contribution of motion in one timescale is not dependent on motion in other timescales. Thus we propose to consider A values rather than S^2 values to compare MD simulation results with NMR relaxation data.

The expressions for $J(\omega)$ and $C(t)$ in the case of extended model-free approach (Eqs. (6) and (3)) differ from those in the case of simple model-free approach (Eqs. (5) and (2)) only by first term corresponding to fast motion. When the first term in (6) cannot be obtained from NMR relaxation data (because it is too small and close to experimental error) it is neglected, i.e. simple model-free approach is applied [2,7]. This occurs when $\tau_f < \tau_s$ (according to our calculations τ_f is about 100–1000 times smaller than τ_s). However, in this case the first term in Eq. (3) cannot be neglected. If $\tau_s < \tau_R$ (for HIV-I protease $\tau_R = 12.1$ ns [2] or more than 100 times as large as τ_s for almost all residues) then the last term in Eqs. (5)–(6) is neglibly small. Under these conditions $J(\omega)$ depends only on the first term in Eq. (5) or on the second term in Eq. (6), i.e. it depends only on values characterizing the slow reorientational motions (with exceptions of overall rotation). If the relaxation data are processed using the simple model-free approach, CMF decay caused by slow motions (i.e. A_s, magnitude of slow motions) obtained from NMR relaxation data is equal to A in Eq. (5) and is expressed as $A_s = 1 - S^2_{eff}$, where S^2_{eff} is S^2 value obtained from MNR relaxation data. In other words, when the extended model-free approach cannot be applied and S^2_f and S^2 values cannot be derived from NMR experiment, one obtains not S^2 but S^2_{eff}. On the other hand, value A_s obtained from MD simulation, is equal to $S^2_f - S^2$ which, in turn, are calculated by smoothing internal CMF. Combining these expressions for A_s we can derive the formula for effective S^2 value which should be compared with experimental values evaluated using the simple model-free approach:

$$S^2_{eff} = 1 + S^2 - S^2_f. \quad (7)$$

Therefore, if NMR relaxation data have been processed with the simple model-free approach, we suggest calculating S^2_{eff} value from MD simulation according to Eq. (7) and comparing it with NMR derived order parameter. We calculated S^2_{eff} values for HIV-1 protease (Fig. 1a). In order to validate our approach we also performed MD simulation for hen lysozyme in water at 300 K. All the MD simulation parameters and procedures were followed after [8]. From 5 ns trajectory of hen lysozyme S^2_{eff} values were computed (Fig. 1b). One can see from Fig. 1 the calculated S^2_{eff} values are in good agreement (comparing to previous S^2 calculations [3–5,8]) with those of simple model-free NMR values and no systematic underestimation is observed. For single residues (especially in loop regions of HIV-1 protease) these parameters are underestimated. We explain this by conformational changes which are observed in MD simulation but occur seldom in reality. Good agreement argues for the correctness of our approach to S^2 calculation from MD simulation data.

4. Conclusion

To compare NMR relaxation data with MD simulation results we propose to use A values from Eqs. (2) and (3) rather than S^2 values. For each kind of motion (motion in each timescale) the proper A value should be evaluated. If S^2 values are obtained from NMR relaxation data using the simple model-free approach, we propose to compare them with effective S^2_{eff} values obtained from MD simulation data using expression (7).

Acknowledgements

We are grateful to “Entry” company for cluster access where MD simulation was carried out. Also we thank Ph.D. Olga Korniushina for assistance in paper preparing.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bpc.2006.03.001.

References

[1] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 1. Theory and range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559.

[2] D.I. Freedberg, R. Ishima, J. Jacobs, et al., Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations, Protein Sci. 11 (2002) 221–232.

[3] I. Chandrasekhar, G.M. Clore, A. Szabo, A.M. Gronenborn, B.R. Brooks, A 500 ps molecular dynamics simulation study of interleukin-1β in water, J. Mol. Biol. 226 (1992) 239–250.
[4] D.B. Kovalskyy, O.S. Ivanova, V.M. Dubyna, D.S. Kanibolotsky, O.I. Kornelyuk, Generalized ordering parameter s2 as measure of the conformational flexibility of proteins: comparison of algorithms of s2 calculation from molecular dynamics simulation data, Ukr. Biochim. Zh. 76 (2004) 128–132.
[5] L. Smith, A. Mark, C. Dobson, W. van Gunsteren, Comparison of MD simulations and NMR experiments for hen lysozyme. Analysis of local fluctuations, cooperative motions, and global changes, Biochemistry 34 (1995) 10918–10931.
[6] E. Lindahl, B. Hess, D. van der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317.
[7] G.M. Clore, P.C. Driscoll, P.T. Wingfield, A.M. Gronenborn, Analysis of the backbone dynamics of interleukin-1β using two-dimensional inverse detected heteronuclear ^{15}N–^{1}H NMR spectroscopy, Biochemistry 29 (1990) 7387–7401.
[8] M. Buck, J. Boyd, C. Redfield, D.A. MacKenzie, D.J. Jeenes, D.B. Archer, C.M. Dobson, Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme, Biochemistry 34 (1995) 4041–4055.